基于 MAC-LSTM 的问题分类研究 *

余本功 [a,b]，许庆堂 [a]，张培行 [a]

(合肥工业大学 a. 管理学院; b. 过程优化与智能决策教育部重点实验室，合肥 230009)

摘　要：问题分类是问答系统问题分析研究的基础组成部分，其精度直接影响自然语言理解效果的好坏。针对问句文本通常较短、语义信息与词语共现信息不足等问题，提出一种多层级注意力卷积长短时记忆模型（multi-level attention convolution LSTM neural network ，MAC-LSTM）的问题分类方法。相比基于词嵌入的深度学习模型，该方法使用疑问词注意力机制对问句中的疑问词特征重点关注。同时，使用注意力机制结合卷积神经网络与长短时记忆模型各自文本建模的优势，既能够并行方式提取词汇级特征，又能够学习更高级别的长距离依赖特征。实验表明，该方法较传统的机器学习方法和普通的卷积神经网络、长短时记忆模型有明显的效果提升。

关键词：问答系统；问题分类；注意力机制；疑问词注意力机制；卷积神经网络；长短时记忆模型

中图分类号：TP391　　doi: 10.19734/j.issn.1001-3695.2018.05.0452

Question classification based on MAC-LSTM

Yu Bengong[a, b], Xu Qingtang[a], Zhang Peihang[a]

(*a. School of Management*, *b. Key Laboratory of Process Optimization & Intelligent Decision-making of Ministry of Education*, *Hefei University of Technology*, *Hefei* 230009, *China*)

Abstract: Question classification is the basic part of the research on question answering system. Its accuracy directly affects the quality of natural language understanding. Most of the question classification methods are based on supervised learning algorithms which require word embedding and does not consider the interrogative features. However, question text is usually short and the information of semantic information and word co-occurrence are not enough. To address the above problems, this paper proposes a multi-level attention convolution LSTM neural network (MAC-LSTM) for question classification. This approach uses the interrogative word attention mechanism to focus on the interrogative features in the heterogeneous question contexts. At the same time, using the attention mechanism combined with the advantages of convolutional neural network and long-short memory model recurrent neural network (LSTM) . MAC-LSTM is able to capture both local features of phrases as well as global and temporal sentence semantics. Experiments show that, our approach achieves better performance than traditional machine learning method, ordinary convolutional neural network, and traditional LSTM on question classification tasks without the need of prior knowledge.

Key words: question answering; question classification; attention mechanism; interrogative attention mechanism; convolutional neural networks; LSTM

0　引言

问答系统(question and answering, QA)是信息检索领域的热门研究领域之一，能够为用户提出的自然语言问题提供一个简明、准确的答案，较好地满足了用户快速、精准地获取信息的需求[1]。问答系统的处理流程主要包括四步，即问题分类、语义理解、文本检索、答案抽取[2]。其中，问题分类通过确定问题的目标答案类型，可以为后续信息检索和答案抽取提供语义限制和约束，缩小候选答案的查找范围。例如，“最有名的徽菜馆在哪里？”是一个关于地点的问题，后续答案抽取时只需要匹配候选答案中的地点类型实体即可，准确有效地提升了问答精度[3]。与此同时，问题分类还可以为问答模式的选择策略提供依据。例如，“正宗徽菜的特点是什么？”是一个描述类问题，生成答案的时候应该侧重于对实体特征的描述与介绍，提供给用户最想获取的答案知识[4]。

与一般的文本分类不同，问题分类的语料具有以下特征：a）问句中的疑问词与问题类别有很强的相关性，加入疑问词特征可以使分类结果更准确；b）问句文本通常较短、特征共现信

收稿日期：2018-05-21；**修回日期：**2018-07-11　**基金项目：**国家自然科学基金资助项目（71671057）

作者简介：余本功（1971-），男，安徽合肥人，教授，博士，主要研究方向为信息系统、机器学习；许庆堂（1994-），男，山东菏泽人，硕士，主要研究方向为深度学习、自然语言处理（xqtbox@163.com）；张培行（1994-），男，山东枣庄人，硕士，主要研究方向为数据挖掘、自然语言处理.

息不足，且具有口语化、歧义性特征。现有的问题分类方法主要是基于规则的方法和基于机器学习的方法。这些方法都没有考虑问题文本语料的特殊性，所以在语义层次上会面临数据特征稀疏和语义敏感问题，且抽取疑问词特征困难。

针对以上问题，本文将深度学习模型与注意力机制相结合，提出一种多层次注意力卷积长短时记忆模型。该方法使用疑问词注意力机制重点关注问句中的疑问词特征，强化了模型的疑问词特征抽取能力。同时，在卷积层与 LSTM 层之间加入注意力机制，使模型既能够并行方式提取词汇级特征，又能够学习更高级别的长距离依赖特征。

1 相关工作

早期的问答系统面向特定领域，强调特定的领域知识，所以问题分类规则需要拥有相关背景的专家制定。比如 Biswas 等人[5]针对医疗数据抽取固定的语法模式，然后利用语法模式来判定问题类别。这种手动设计的特征规则需要大量人力，并且只适用于特定的数据集，可迁移性差。而基于机器学习的问题分类方法具有较强的适用性，这种分类方法的研究主要把精力放在机器学习模型的选择和特征提取上面，即如何在较短的问题文本上面抽取特征、提取出浅层的语义信息。比如 Li 等人[6]利用双语（英语汉语）、句子长度、句子音节数等信息作为支持向量机模型的扩展特征，对英文句子进行分类并取得了优良效果。考虑到问句中疑问词的重要性，Hu 等人[7]提出了疑问语义单元的概念，并结合 HOWNET 对其进行语义扩展，在开放式问题分类实验中达到了较高的分类精确度。但是基于机器学习的问题分类算法面临着数据特征稀疏的问题，不能很好地抽取文本中蕴含的语义信息，分类器的效果仍然有待提升。

当前，深度学习在自然语言处理等领域取得了优秀表现，也有许多学者利用深度学习技术[8]来解决问答系统问题分类任务。Li 等人[9]将卷积神经网络应用于知识库问答系统问题分类，并扩展了答案文本、答案类型、答案路径等特征来学习问题的分类规则。对于同样的问题，Feng 等人[10]使用共享卷积神经网络来训练问题答案对，并在此基础上进行了语义相似度计算。这些方法使用了卷积神经网络来解决问题分类任务，能够较好地捕捉从数据本身到高层语义的复杂映射，具有远超于传统机器学习模型的表达能力。但是这些方法的参数复杂且没有考虑文本语料的时序特征。针对这一问题，有一些学者开始使用长短时记忆模型来处理问句这种序列数据。长短时记忆模型是一种特殊的循环神经网络改进结构，能够解决原模型梯度消失和长距离依赖的问题[11]。Wang 等人[12]基于多层长短时记忆网络模型训练问答对的联合特征向量，把问答匹配问题转换为分类或排序问题。此外，考虑到卷积神经网络和长短时记忆模型各自对文本建模的优点，Zhou 等人[13]将两种模型相结合提出了一种新颖的组合模型（C-LSTM），并指出卷积神经网络和长短时记忆模型提供了互补的信息，该模型在情感分类和问题分类两个任务中都取得了优良的表现。

注意力机制的原理是仿照人脑的注意力思维，通过自动加权的方式对整体信息进行局部重点关注，帮助深度学习模型取得显著的精度提升[14,15]。2014 年，Bahdanau 等人[16]使用基于注意力机制的循环神经网络来解决机器翻译任务，开启了注意力机制在自然语言处理领域的应用。Zhou 等人[17]提出了双语长短时记忆模型的层次注意力机制来解决跨语言情感分析任务，使用单词级注意力模式可以知道每个句子中的哪个词是具有决定性的，而句子级注意力模式学习哪些句子对于确定整体文档的情绪更为重要。另外，考虑到卷积神经网络在短文本分类场景中的优良效果，有学者探究了注意力机制与卷积神经网络的结合[18]。其中，文献[19]通过注意力机制的方法将用户和产品信息与文本语义信息结合，提出了一个层次化卷积神经网络。该方法能够在一定程度减少模型参数，并使得文本中语义信息更加丰富。

以上研究成果促使人们从问题分类领域引用注意力机制来衡量特征的重要性，结合相关模型可以实现特征的有效提取。本文提出的多层级注意力卷积长短时记忆模型利用卷积神经网络对疑问词注意力矩阵提取局部特征，并结合注意力矩阵筛选对问题分类最相关的特征映射输入给双向长短时记忆模型。此外，在问答领域的问题分类任务中，疑问词是具有重要影响的特征。为了有效提取疑问词特征，本文提出了一种基于疑问词的注意力机制，通过构造疑问词注意力矩阵强化模型对疑问词特征的重点关注。这样 MAC-LSTM 同时拥有长短时记忆模型捕获全局上下文和卷积神经网络捕获问句局部特征的优势，并拥有了对疑问词等重点特征的抽取能力。

2 基于 MAC-LSTM 的问题分类

本文提出的 MAC-LSTM 模型通过两个层次的注意力机制学习问题文本的疑问词特征，有效提高问题分类的准确度。首先，在模型输入层引入疑问词注意力机制，通过疑问词典构建疑问词注意力矩阵，强化问句表示中的疑问词语义信息，并将疑问词注意力矩阵输入卷积层提取局部特征；其次，模型使用注意力矩阵来筛选对问题分类最有用的卷积特征，输送给双向长短时记忆模型层进行高层次时序特征抽取；最后，将融合后的特征向量加入到 Softmax 分类器中完成问题分类任务。模型结构如图 1 所示。

2.1 基于疑问词注意力机制的输入表示

与普通文本分类不同，问句的类型判别更加依赖于疑问词这一分类特征。这是由于问题文本较短、语义信息和词语共现信息不足，所以问句中的疑问词对问题分类的结果影响较大。而深度学习模型一般使用的词向量并没有对疑问词信息进行重点关注。针对该问题，本文提出了一种基于疑问词的注意力机制，通过构建疑问词典，将词向量与疑问词对角关注矩阵组合，使模型有能力确定哪一部分句子对疑问词最有影响力。具体方法为：收集汉语问句中的疑问词建立疑问词典，如“谁”“哪里”“多少”等都属于疑问词；然后对问句分词后查找疑问词典，

找到问句中的疑问词 e；随后以词为单位训练词向量 W_n；最后引入对角关注矩阵 A^r 来表征句子中普通词语 $w_i \in W_n$ 与疑问词 e 之间的语境相关性和连接强度。其元素 A_i^r 的计算公式如下：

$$A_i^r = f(e, w_i) \cdot \beta \tag{1}$$

其中：疑问词 e 通过本文构造的疑问词典，扫描问句分词序列得到；函数 f 为词向量的内积运算；β 为参数向量，并且在训练过程期间通过反向传播被新。同时对于每一个对角关注矩阵元素 A_i^r，定义：

$$\alpha_i^r = \frac{\exp(A_i^r)}{\sum_{j=1}^{n} \exp(A_j^r)} \tag{2}$$

来表示第 i 个普通词 w_i 与疑问词 e 之间的相对重要程度，并在此基础上构造基于注意力机制的疑问词向量：

$$\mathrm{x}_i = \alpha_i^r \cdot \mathrm{w}_i \tag{3}$$

模型输入表示层的最终结果是基于注意力机制的疑问词向量排列矩阵：

$$\mathrm{X} = [\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n] \tag{4}$$

其中：矩阵维度为 $R^{l \times n}$， l 为词向量的维度，n 是问句长度；逗号表示行矢量级连接。图 2 描述了基于注意力机制的疑问词向量表示矩阵构造的详细过程。

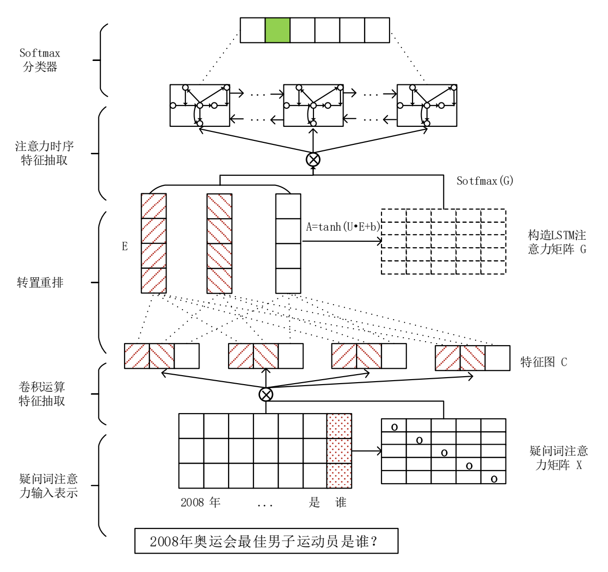


图 1　基于 MAC-LSTM 的问题分类模型

Fig.1　Question classification model based on MAC-LSTM

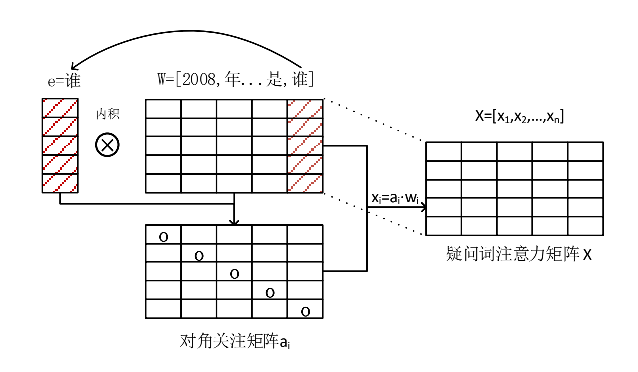


图 2　疑问词向量注意力矩阵构建

Fig.2　Interrogative word embedding attention matrix

2.2　通过卷积运算进行特征抽取

卷积神经网络能够通过卷积运算从原始数据中抽取局部特征。本文使用 d 个相同大小的一维滤波器在疑问词向量序列上滑动，以提取不同位置的文本特征，得到疑问词语义依存关系特征映射。一维滤波器表示为 $m \in R^{l \times k}$，其中 k 为滤波器的宽度。此时，句子中的第 i 个词语对应的窗口矩阵 w_i 由 k 个一维向量组成：

$$\mathrm{W}_i = [x_i, x_{i+1}, \cdots, x_{i+k-1}] \tag{5}$$

其中：逗号表示行矢量级连接。卷积滤波器 m 与每个位置处的窗口矩阵(k-gram)卷积运算生成特征映射 $\mathrm{c} \in \mathrm{R}^{n-k+1}$，这里特征映射每个元素 c_j 的计算公式定义如下：

$$\mathrm{c}_j = f(\mathrm{W}_i * m + b) \tag{6}$$

其中：$*$ 为元素乘法；b 为偏差项。相关研究表明 ReLU 具有单侧抑制性和相对宽阔的兴奋边界，并通过引入网络稀疏性从而获得更快的特征学习速率[20]。本文非线性激活函数 f 选择 ReLU，公式如式(7)所示。

$$f(x) = max(0, x) \tag{7}$$

为保留原文数据的时序性特征，对于 d 个具有相同长度的特征映射 c，将其转置重排为对应与每个窗口矩阵 W_i 的特征表示：

$$\mathrm{E} = [c_1; c_2; \cdots; c_d] \tag{8}$$

其中：分号表示列向量的连接；c_i 是使用第 i 个滤波器生成的特征向量。 $\mathrm{E} \in \mathrm{R}^{(d) \times (n-k+1)}$ 的每列 E_j 是原问题文本位置 j 处窗口向量的新特征表示形式。

2.3　基于注意力机制的时序特征抽取

对于问题分类任务，问题的分类规则依赖于前向和后向的上下文信息，因此本文将卷积层提取的窗口映射输入给双向长短时记忆模型进一步提取文本时序特征。双向长短时记忆模型层结构如图 3 所示。

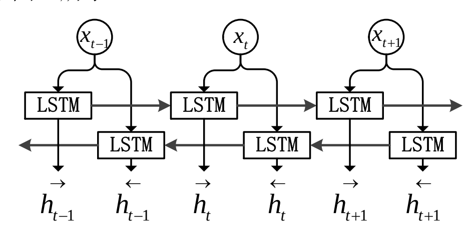


图 3　双向长短时记忆模型结构

Fig.3　Bidirectional LSTM

图 3 中，方框表示长短时记忆单元；$\vec{h}$ 为前向长短时记忆模型在 t 时刻的输出；$\overleftarrow{h}$ 为后向长短时记忆模型在 t 时刻的输出。通过串接操作将前向序列和后向序列的输出进行合并，即可得到文本的上下文语义表示。

为了让整体模型在句子层面有能力自动识别句子中对问题分类最相关的部分，本文提出了一种基于注意力机制的连接模式来连接卷积层与长短时记忆模型。具体做法为：对卷积特征构造加权注意力矩阵 G_i 来比较卷积特征的重要性。通过不同的权值大小反映语义重要性的大小，对更重要的句子部分进行重点关注，同时减少了特征提取过程中的信息丢失和信息冗余。

注意力矩阵计算方式如下：

$$A_i^c = tanh\left(U \cdot E_j + b\right) \tag{9}$$

其中：U 是神经网络学习的加权参数矩阵；b 表示注意力机制的线性偏差。A_i^c 为 E_j 对问题分类的重要程度函数，进一步对其归一化处理：

$$a_i^c = \frac{\exp\left(A_i^c\right)}{\sum_{i=1}^{m}\exp\left(A_i^c\right)} \tag{10}$$

最后，将这个注意力矩阵与卷积层的输出 E_j 相乘，把向量 G_i 表示对分类结果重要的部分并作为双向长短时记忆模型的输入。

$$G_i = a_i^c \cdot E_j \tag{11}$$

经过双向长短时记忆模型的时序语义建模，最后将隐藏层的输出向量作为 Softmax 分类器的输入，从全局角度对特征进行分析，进而完成问题模型的分类任务。

3　实验与分析

为了验证本文模型的有效性，解决问题分类任务，本文将在不同的中文短文本数据集上进行对比实验。实验环境为如表 1 所示。

为进行对比实验，本文采用与 Zhou 等人[13]相同的词向量训练配置，使用 word2vec 的 skip-gram 训练模式，上下文窗口大小设置为 5，词向量维度大小设为 100，得到原始短文本的输入表示即词组映射矩阵 。

表 1　实验环境及配置

Table 1　Experimental environment and configuration

实验环境	环境配置
操作系统	Windows10 企业版
CPU	Inter Core i5-6500 3.2GHz
GPU	NVIDIA GeForce GTX 1070
内存	16G
编程语言	Java
分词工具	ICTCLAS 2016
深度学习框架	DeepLearning4J

此外，问题分类体系的构建是问题分类的前提[21]。对于英文问题分类，各机构往往对 TREC QA 标准数据集采用 UIUC 分类体系，将问题分为 ABBR、DESC、ENTY、HUM、LOC 和 NUM 六大类。而中文问题分类没有统一的分类体系，被大多数学者采用的是哈尔滨工业大学信息检索和社会计算中心所提出的分类体系，它根据汉语的特点提出包括描述类（DES）、人物类（HUM）、地点类（LOC）、数字类（NUM）、时间类（TIME）、实体类（OBJ）六个大类，每个大类及其实例如表 2 所示。

3.1　数据集的选择与处理

本文选用的数据集分为三个部分：百度实验室数据集 6 205 条数据①，CCF 国际自然语言处理与中文计算会议 2016（NLPCC 2016）QA 评测的问题集共 9 604 条数据②，NLPCC 2017 QA 评测的问题集共 9 518 条数据③，共有 25 327 条数据。其中 NLPCC QA 提供的公开数据集格式清晰、质量高，但数据集中只有问题与答案对，所以本文对该数据进行了人工标注工作。在本任务中，每个问题都由 3 人独立标注，对于有异议的数据存储数据库最后协商标注。同时，为排除数据集里偶然误差的干扰，本文将各数据集大致按照 20%的比例随机划分训练集与测试集。

表 2　分类体系

Table 2　Category system

大类	代表例子
描述类（DES）	万达广场主要经营的是什么生意？
人物类（HUM）	《机械设计基础》这本书的作者是谁？
地点类（LOC）	安德烈是哪个国家的人呢？
数字类（NUM）	合肥地铁一号线总投资额是多少？
时间类（TIME）	华严寺何时建造的？
实体类（OBJ）	中国第二大民族是什么族？

3.2　对比实验设置

本文提出的 MAC-LSTM 模型通过多层级注意力机制将卷积神经网络与长短时记忆模型相结合，针对问题分类任务对疑问词特征重点关注。为验证模型的有效性，本文设置多种模型方法在两种数据集上进行对比实验.

a）SVM。基于 Li 等人[6]提出的使用线性核函数的 SVM 模型，采用词袋模型进行文本表示，并运用 TF-IDF 算法对单词进行权重计算，是效果较好的传统分类模型。

b）CNN。由 Kim[8]提出的基础卷积神经网络模型，由卷积层、池化层以及全连接层组成。

c）LSTM。文献[10]提出的一种双向长短时记忆模型，适合于处理和预测时间序列中间隔和延迟相对较长的文本序列。

d）C-LSTM。Zhou 等人[13]将卷积神经网络与长短时记忆模型相结合，由卷积层提出特征后输入给长短时记忆模型，采用了新颖的向量重排模式。

e）MCA-LSTM。本文提出的模型，在 C-LSTM 的基础上增加了基于注意力机制的疑问词矩阵输入和连接层注意力矩阵，能够自适应的识别句子中最重要的部分。

3.3　实验结果与分析

通过与不同模型对比来验证 MAC-LSTM 模型在短文本分类中的适用性和优越性，本文在不同数据集进行了对比实验。实验结果如表 3 所示。

表 3　不同模型准确率对比

Table 3　Comparison of evaluation scores

模型	Bidu
SVM	76.69%
CNN	89.14%
LSTM	90.82%
C-LSTM	92.65%
MAC-LSTM	94.31%

从表 3 来看，由于不同数据集各大类数据的数量分布不同，各模型精确度有不同程度的波动，但并不影响各模型对比参照。本文画出直方图直观的展示几类模型在不同数据集上的精度对比，MAC-LSTM 在三个数据集上都表现出较好的优势。

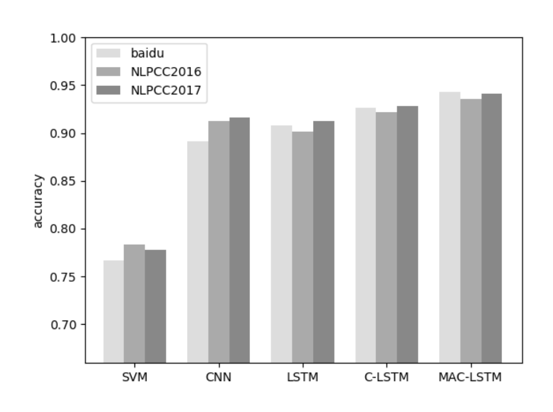


图 4　模型三个数据集上的精度对比

Fig.4　Comparison of different database

图 4 展示了本文的模型结果优于高度人工设计特征的传统模型 SVM，人工设计特征需要大量的人力劳动，不能很好地推广其他数据集和任务。本文提出的 MAC-LSTM 模型具有自动学习语义句子表示的能力，不需要任何人工抽取特征具有更好的可扩展性；将单个卷积神经网络和长短时记忆模型的结果与相比，卷积神经网络在 NLPCC 数据集精确度更高，表明在文本较短且数据量充足的情况下卷积运算更能有效抽取文本特征；将 C-LSTM 与本文提出的 MAC-LSTM 与单一的卷积神经网络和长短时记忆模型相比，分类效果更好，说明模型的结合在抽取文本局部特征的同时，可对序列特征进一步深度挖掘，更有利于对问题的理解；对比本文的 MAC-LSTM 模型与 C-LSTM 方法可以看出，本文提出的多层级注意力机制能使模型在训练过程高度关注分类的目标的特征信息，从而更好地识别疑问词相关特征，验证了注意力机制在问题分类任务中的有效性。

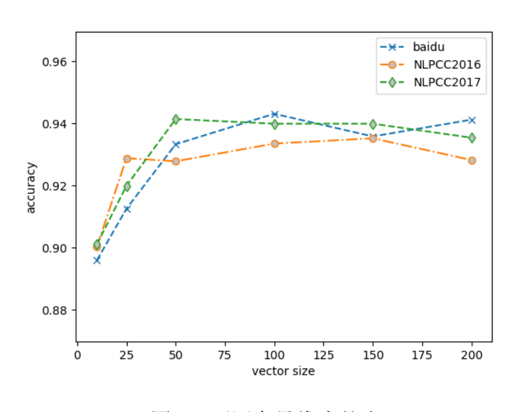


图 5　不同向量维度的表现

Fig.5　Comparison of different embedding dimensions

本文在 C-LSTM 研究的基础上提出了疑问词注意力机制，通过构建疑问词向量强化模型对问句疑问词特征的抽取效果。疑问词向量的特征表示效果与词向量的维度有很大关系，本文使用不同维度的词向量构造疑问词向量进行实验，验证疑问词注意力机制的有效性和稳定性。实验结果如图 5 所示。从图中可以看出，当向量维度大小变化时，模型准确率在开始时会增加；当维度大小超过 50 之后变得稳定。本文最终选择 100 作为向量维度，在兼顾到计算简单的情况下能得到相对好的实验结果。

4　结束语

本文从问题分类的角度出发，分析了传统方法的不足之处以及深度学习在本领域相关的应用，同时结合了注意力机制对深度学习模型进行改进，本文提出了一种多层次注意力深度学习模型。首先应用注意力机制加强了模型对疑问词的特征的关注度；然后将卷积神经网络与长短时记忆模型通过注意力连接模式有效结合，同时发挥两类模型对问题分类的优势。MAC-LSTM 模型能够通过卷积运算学习疑问词向量的短语级特征，然后将特征表示输送到长短时记忆模型中以增强捕捉文本时序依赖特征的能力。本文使用 MAC-LSTM 模型对问题类型分类任务进行了评估实验，取得了满意的结果。当前问题分类任务面临着缺少语料的问题，除了继续搜集和标注新的数据这个思路之外，如何利用大量未标记的语料进行半监督学习也是一条可行的方向。此外，问题分类是问答系统的基础研究之一，问题分类结果的好坏会影响问答系统答案抽取与回答生成策略地选择，所以答案抽取与问题分类相结合方面的研究也将是下一步研究的重点。

参考文献：

[1] Dang H T, Kelly D, Lin J. Overview of the TREC 2007 question answering track [C]// Proc of the 16th Text Retrieval Conference. 2007: 115-123.

[2] Le-Hong P, Phan X H, Nguyen T D. Using dependency analysis to improve question classification [C]// Proc of Knowledge and Systems Engineering. 2015: 653-665.

[3] Hernault H, Prendinger H, Duverle D A, *et al.* HILDA: a discourse parser using support vector machine classification [J]. Dialogue & Discourse, 2010, 1 (3) .

[4] Li Xin, Roth D. Learning question classifiers [J]. Proc. COLING-2002, 2002, 12 (24): 556-562.

[5] Biswas P, Sharan A, Kumar R. Question classification using syntactic and rule based approach [C]// Proc of International Conference on Advances in Computing, Communications and Informatics. 2014: 1033-1038.

[6] Li Kuan, Xiong Zhongyang, Zhang Yufang, *et al.* The spoken//written language classification of english sentences with bilingual information [M]// Natural Language Processing and Chinese Computing. Berlin: Springer, 2013: 370-377.

[7] Hu Yue, Liu Bo, Zhang Shouwei. Chinese question classify model based on interrogative semantic unit [C]// Proc of International Symposium on Computer Architecture. 2013, 92: 14-19.

[8] Kim Y. Convolutional neural networks for sentence classification [J]. arXiv



preprint arXiv: 1408. 5882, 2014.

[9] Li Dong, Wei Furu, Zhou Ming, *et al.* Question answering over freebase with multi-column convolutional neural networks [C]// Proc of Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing. 2015: 260-269.

[10] Feng Minwei, Xiang Bing, Glass M R, *et al.* Applying deep learning to answer selection: a study and an open task [J]. Astrophysics Data System, 2015: 813-820.

[11] Ruder S, Ghaffari P, Breslin J G. A hierarchical model of reviews for aspect-based sentiment analysis [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2016: 999-1005.

[12] Wang Ding, Nyberg E. A long short-term memory model for answer sentence selection in question answering [C]// Proc of Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing. 2015: 707-712.

[13] Zhou Chunting, Sun Chonglin, Liu Zhiyan, *et al.* A C-LSTM neural network for text classification [J]. arXiv preprint arXiv: 1511. 08630, 2015.

[14] Yin Wenpeng, Schütze H, Xiang Bing, *et al.* Abcnn: attention-based convolutional neural network for modeling sentence pairs [J]. arXiv preprint arXiv: 1512. 05193, 2015.

[15] Mnih V, Heess N, Graves A, *et al.* Recurrent models of visual attention [J]. Astrophysics Data System, 2014, 3: 2204-2212.

[16] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. arXiv preprint arXiv: 1409. 0473, 2014.

[17] Zhou Xinjie, Wan Xiaojun, Xiao Jianguo. Attention-based LSTM network for cross-Lingual sentiment classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2016: 247-256.

[18] Yang Zichao, Yang Diyi, Dyer C, *et al.* Hierarchical attention networks for document classification [C]// Proc of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2017: 1480-1489.

[19] Chen Huimin, Sun Maosong, Tu Cunchao, *et al.* Neural sentiment classification with user and product attention [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2016: 1650-1659.

[20] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines [C]// Proc of International Conference on International Conference on Machine Learning. 2010: 807-814.

[21] Hao Tianyong, Xie Wenxiu, Wu Qingyao, *et al.* Leveraging question target word features through semantic relation expansion for answer type classification [J]. Knowledge-Based Systems, 2017.